UNITED STATES PATENT AND TRADEMARK OFFICE

CORRECTED
NOTICE OF ALLOWANCE AND FEE(S) DUE

62296   7590   04/03/2008

GOOGLE / FENWICK
SILICON VALLEY CENTER
801 CALIFORNIA ST.
MOUNTAIN VIEW, CA 94041

EXAMINER
HERNANDEZ, JOSIAH J

ART UNIT	PAPER NUMBER
2626	

DATE MAILED: 04/03/2008

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/634,616	08/04/2003	Alexander Franz	025.0289.US.UTL	6590

TITLE OF INVENTION: IDENTIFYING LANGUAGE ATTRIBUTES THROUGH PROBABILISTIC ANALYSIS

APPLN. TYPE	SMALL ENTITY	ISSUE FEE	PUBLICATION FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	NO	$1440	$0	$1440	05/19/2008

THE APPLICATION IDENTIFIED ABOVE HAS BEEN EXAMINED AND IS ALLOWED FOR ISSUANCE AS A PATENT. PROSECUTION ON THE MERITS IS CLOSED. NEITHER A NOTICE OF ALLOWANCE NOR A CORRECTED NOTICE OF ALLOWANCE IS A GRANT OF PATENT RIGHTS. THIS APPLICATION IS SUBJECT TO WITHDRAWAL FROM ISSUE AT THE INITIATIVE OF THE OFFICE OR UPON PETITION BY THE APPLICANT. SEE 37 CFR 1.313 AND MPEP 1308.

THE ISSUE FEE AND ANY PUBLICATION FEE (IF REQUIRED) MUST BE PAID WITHIN THE THREE MONTH PERIOD BEGINNING ON THE MAILING DATE OF THE PREVIOUSLY-MAILED NOTICE OF ALLOWANCE AND ENDING ON THE DATE DUE SHOWN ON THIS FORM, OR THIS APPLICATION SHALL BE REGARDED AS ABANDONED. THIS STATUTORY PERIOD CANNOT BE EXTENDED. SEE 35 U.S.C. 151. MAILING OF THIS CORRECTED NOTICE OF ALLOWANCE DOES NOT CHANGE THE DATE DUE OF THE ISSUE FEE (AND ANY REQUIRED PUBLICATION FEE). IF A REPLY (WITH PAYMENT OF THE ISSUE FEE AND ANY PUBLICATION FEE) WAS FILED IN RESPONSE TO THE PREVIOUSLY-MAILED NOTICE OF ALLOWANCE, THEN NO FURTHER REPLY IS REQUIRED FROM APPLICANT.

All communications regarding this application must include the application number. Please direct all communications prior to issuance to Mail Stop ISSUE FEE, unless advised to the contrary.

UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/634,616	08/04/2003	Alexander Franz	025.0289.US.UTL	6590

				EXAMINER
62296	7590	04/03/2008		HERNANDEZ, JOSIAH J

GOOGLE / FENWICK
SILICON VALLEY CENTER
801 CALIFORNIA ST.
MOUNTAIN VIEW, CA 94041

ART UNIT	PAPER NUMBER
2626	

DATE MAILED: 04/03/2008

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(application filed on or after May 29, 2000)

The Patent Term Adjustment to date is 805 day(s). If the issue fee is paid on the date that is three months after the mailing date of this notice and the patent issues on the Tuesday before the date that is 28 weeks (six and a half months) after the mailing date of this notice, the Patent Term Adjustment will be 805 day(s).

If a Continued Prosecution Application (CPA) was filed in the above-identified application, the filing date that determines Patent Term Adjustment is the filing date of the most recent CPA.

Applicant will be able to obtain more detailed information by accessing the Patent Application Information Retrieval (PAIR) WEB site (http://pair.uspto.gov).

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571) 272-7702. Questions relating to issue and publication fee payments should be directed to the Customer Service Center of the Office of Patent Publication at 1-(888)-786-0101 or (571)-272-4200.

PTOL-85 (Rev. 08/07) Approved for use through 08/31/2010.

	Application No.	Applicant(s)
supplemental ***Notice of Allowability***	10/634,616	FRANZ ET AL.
	Examiner	Art Unit	
	JOSIAH HERNANDEZ	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☒ This communication is responsive to _08/04/2003_.

2. ☒ The allowed claim(s) is/are _1,3-10,14,15,17-24,28-30,32,33 and 35-37_.

3. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a) ☐ All   b) ☐ Some*   c) ☐ None   of the:

        1. ☐ Certified copies of the priority documents have been received.

        2. ☐ Certified copies of the priority documents have been received in Application No. _____ .

        3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the

            International Bureau (PCT Rule 17.2(a)).

    * Certified copies not received: _____.

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application. **THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.**

4. ☐ A SUBSTITUTE OATH OR DECLARATION must be submitted. Note the attached EXAMINER'S AMENDMENT or NOTICE OF INFORMAL PATENT APPLICATION (PTO-152) which gives reason(s) why the oath or declaration is deficient.

5. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.

    (a) ☐ including changes required by the Notice of Draftsperson's Patent Drawing Review (PTO-948) attached

        1) ☐ hereto or 2) ☐ to Paper No./Mail Date _____.

    (b) ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of

        Paper No./Mail Date _____.

    Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).

6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. ☐ Notice of References Cited (PTO-892)

2. ☐ Notice of Draftperson's Patent Drawing Review (PTO-948)

3. ☐ Information Disclosure Statements (PTO/SB/08), Paper No./Mail Date _____

4. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material

5. ☐ Notice of Informal Patent Application

6. ☐ Interview Summary (PTO-413), Paper No./Mail Date _____ .

7. ☐ Examiner's Amendment/Comment

8. ☒ Examiner's Statement of Reasons for Allowance

9. ☐ Other _____.

Detailed Action

Response to Amendment

1.   In response to the office Action filed August 7, 2007, applicant has submitted an

Amendment, filed November 8, 2007 canceling claims 2, 11-13, 16, 25-27, 31, and 34.

Applicant has also amended independent claims 1, 15, 30, 33 and 37, reciting

calculating an overall probability for ones of the set of languages classes by evaluating

the probability for ones of the set of language classes by evaluating the probability for

the document properties set based on the attribute model and the probability for the

byte occurrences based on the text model.

EXAMINER'S AMENDMENT

2.   An examiner's amendment to the record appears below. Authorization for this

examiner's amendment was given in a telephone interview with Robert R. Sachs on

January 17, 2008.

The Independent claims 1, 15, and 30 have been amended to state the following:

1. A system for identifying language attributes through probabilistic analysis,
comprising:

a storage system adapted to store a set of language classes, which each

identify a language and a character set encoding, and further adapted

to store a plurality of training documents;

an attribute modeler adapted to train an attribute model by evaluating occurrences of one or more document properties within the training documents and, for each language class, calculating a probability for [[the]] a set of the one or more document properties, [[set]] the probability conditioned on the occurrence of the language class, the trained attribute model stored in the storage; wherein the document properties comprise at least one of top level domain, HTTP content character set encoding and language header parameters, and HTML content character set encoding and language metatags;

a text modeler adapted to train a text model by evaluating byte occurrences within the training documents and, for each language class, calculating a probability for [[the]] a set of byte occurrences, the probability conditioned on the occurrence of the language class, the trained text model stored in the storage; and

a training engine adapted to calculate an overall probability for [[ones]] at least one of the set of language classes by evaluating the probability for the document properties set based on the attribute model and the probability for the byte occurrences based on the text model.

15. A method for identifying language attributes through probabilistic analysis, comprising: defining a set of language classes, which each identify a language and a character set encoding, and a plurality of training documents; evaluating occurrences of one or more document properties within the training documents and, for each language class, calculating a probability for the document properties set conditioned on the occurrence of the language class by an attribute model wherein the document properties comprise at least one of top level domain, HTTP content character set encoding and language header parameters, and HTML content character set encoding and language metatags; [[and]]

evaluating byte occurrences within the training documents and, for each language class, calculating a probability for the byte occurrences conditioned on the occurrence of the language class by a text model[[.]]; and calculating an overall probability for ones of the set of language classes by evaluating the probability for the document properties set by the attribute model and the probability for the byte occurrences by the text model.

30. A system for identifying documents by language using probabilistic analysis of language attributes, comprising a set of language classes, each language class comprising a language name and a character set encoding name; a training corpora comprising a plurality of training documents; an attribute modeler adapted to train an attribute model by evaluating a top level domain and character set encoding associated with the training documents and, for each language class, calculating a probability for each such top level domain and character set encoding conditioned on the occurrence of the each language class wherein the document properties comprise at least one of top level domain, HTTP content character set encoding and language header parameters, and HTML content character set encoding and language metatags; [[and]] a text modeler adapted to train a text model by evaluating co-occurrences of a plurality of bytes within the training documents and, for each language class, calculating a probability for the byte co-occurrences conditioned on the occurrence of the each language class[[.]]; and a training- eng- ine adapted to calculate an overall probability for ones of the set of language classes bv evaluating the probability for the top level domain and character set encoding- based on the attribute model and the probability for the byte occurrences based on the text model.

The original dependent claim 32 has been amended to state the following:

32. A system according to Claim <u>30</u>, further comprising: a plurality of unlabeled documents; and a classifier classifying one or more unlabeled documents by at least one language class, comprising: an attribute evaluator determining document properties within the documents and initializing language class probability to each document from the attribute model; a text evaluator evaluating byte occurrences in the documents and updating the language class probability of the each document from the text model; a pruner pruning at least one language class falling below a predetermined probability threshold; and an assignment module assigning at least one language class based on the language class probability of each document.

Allowable Subject Matter

3.   Claims 1, 3-10, 14-15, 17-24, 28-30, 32-33, and 35-37 are allowed.

4.   The following is an examiner's statement of reasons for allowance:

As to claims 1, 15, 30, 33, and 37, there is no prior art reference, alone or in combination, that teaches or fairly suggests evaluating occurrences of one or more

document properties within the training documents and, for each language class,

calculating a probability for the document properties; wherein the document properties

comprise at least one of top level domain, HTTP content character set encoding and

language header parameters, and HTML content character set encoding and language

metatags.

Bracewell et al. (US PGPUB 2006/0041685) teaches that document properties

like HTTP header information can be used in order to identify the language of a

document or search query on the internet (paragraph [0014]). However, none teach or

suggest using the document information such as: top level domain, HTTP content

character set encoding and language header parameters, and HTML content character

set encoding and language metatags, in order to calculate the occurrences of that

specific property and further calculating probabilities from the occurrences of the

properties that can be directly associated to a particular language and be used as part

of a language model in order to identify the language of the document.

Claims 3-10, 14, 17-24, 28, 29, 32, 35, and 36 are allowed because they further

limit their parent claims, which are allowed.

Any comments considered necessary by applicant must be submitted no later

than the payment of the issue fee and, to avoid processing delays, should preferably

accompany the issue fee. Such submissions should be clearly labeled "Comments on

Statement of Reasons for Allowance."

Conclusion

A note has been made to notify the appropriate parties that the examiner has moved

from Art Unit 2609 to 2626.

Any inquiry concerning this communication should be directed to Josiah

Hernandez whose telephone number is 571-270-1646. The examiner can

normally be reached from 7:30 pm to 5:00 pm.

If attempts to reach the examiner by telephone are unsuccessful, the

examiner's supervisor, David Hudspeth can be reached on (571) 272-7843. The

fax phone number for the organization where this application or proceeding is

assigned is 703-872-9306.

Information regarding the status of an application may be obtained from

the Patent Application Information Retrieval (PAIR) system. Status information

for published applications may be obtained from either Private PAIR or Public

PAIR. Status information for unpublished applications is available through

Private PAIR only. For more information about the PAIR system, see http://pair-

direct.uspto.gov. Should you have questions on access to the Private PAIR

system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-

free).

JH

/David R Hudspeth/

Supervisory Patent Examiner, Art Unit 2626